



Diurnal variation of major error rates in the interpretation of abdominal/pelvic CT studies

Mark A. Kliewer¹ · Lu Mao² · Mikala R. Brinkman³ · Richard J. Bruce¹ · J. Louis Hinshaw¹

Received: 21 July 2020 / Revised: 21 September 2020 / Accepted: 30 September 2020 / Published online: 10 October 2020
© Springer Science+Business Media, LLC, part of Springer Nature 2020

Abstract

Rationale and objective Variation of visual selective attention through the day has been demonstrated in several arenas of human performance, including radiology. It is uncertain whether this variation translates to an identifiable diurnal pattern of error rates for radiology interpretation. The purpose of this study was to attempt to identify particular days of the week and times of the day when radiologists might be most prone to error.

Materials and methods Abdomen/pelvis CT studies containing at least one major error were collected from a 10-year period from the quality assurance (QA) database at our institution. A major error was defined as a missed finding that had altered management in a way potentially detrimental to the patient. The identified studies were categorized by the day of the week and hour of the day that the study was interpreted. Study volume data over this same period was also obtained by day of the week and time of day, so to normalize the data based on case volume. Standard errors of the volume-adjusted error rates were obtained based on the binomial distribution. The null hypothesis of constant error rates over time was tested using a weighted logistic regression model with linear time as predictor.

Results A total of 252 major errors were identified. More errors were made on Monday than on any other day of the week ($n = 58$). Major error rates increased through the mid to late morning (9 am to 12 pm), and then decreased progressively through the afternoon until 4 pm, when a rise in the error rate was seen. This pattern persisted when error rates were normalized by study volume within each hour. Overall tests of time-constancy of error rates by day and hour were statistically significant (both p -values < 0.001).

Conclusion Our study shows that error rates in abdominal CT do seem to vary with time of day and day of the week. During the workweek, error rates were highest in the late morning and at the close of the workday, and greater on Mondays than other days.

Keywords Diurnal variation · Error · CT

✉ Mark A. Kliewer
mkliewer@uwhealth.org

Lu Mao
lmao@biostat.wisc.edu

Mikala R. Brinkman
mikalabrinkman@yahoo.com

Richard J. Bruce
RBruce@uwhealth.org

J. Louis Hinshaw
JHinshaw@uwhealth.org

¹ Department of Radiology, University of Wisconsin School of Medicine, E3/311 Clinical Science Center, 600 Highland Avenue, MC 3252, Madison, WI 53792-3252, USA

² Department of Biostatistics and Medical Informatics, University of Wisconsin School of Medicine, 600 Highland Avenue, MC, Madison, WI 53792, USA

³ Central Illinois Radiological Associates, Ltd, 111 Oakwood Road, East Peoria, IL 61611, USA

Introduction

Variation of mental acuity and visual selective attention has been demonstrated in several arenas of human performance, including radiology [1–6]. It is unclear, however, whether such variation follows a diurnal pattern for practicing radiologists or whether it results in periods of the day when the radiologist might be particularly prone to error. That such a pattern might exist is intuitively plausible, particularly for studies that are as mentally taxing and complex as CT and MR studies, many of which consist of hundreds or thousands of images.

Two studies of plain film interpretation have reported declines in performance from the morning to the afternoon [3, 4]. An often-referenced study of chest radiograph interpretation involved medical students looking at chest radiographs, and reported a morning-to-afternoon decline in performance. Medical students have little training or competence in radiology, and therefore these results are difficult to generalize to the practice of radiology by radiologists, and even less so to the practice of subspecialty radiologists. The other study of mammogram interpretation found that recall rates were lower in the afternoon for some readers, and not for all [4]. Another mammography study, conducted during a national meeting, found that detection rates of breast lesions did vary with time of day but that no one time of day was best for all readers [5].

Therefore, the impact of the time of day on radiologists' performance remains an unresolved issue even with plain film interpretation. The current literature is essentially silent with regard to diurnal variation for the interpretation of complex cross-sectional imaging studies [6].

In this study, we examined major error rates in the interpretation of abdominal/pelvic CT cases over a 10-year period and categorized these by time of day and day of the week. The purpose of this study was to attempt to identify particular days of the week and times of the day when radiologists might be most prone to error.

Methods and materials

This study was conducted under the auspices of the quality assurance program at our institution, HIPAA compliant, and exempt from IRB review. Within the radiology department at our institution, a quality assurance database was established and maintained on the PACS system. Cases were entered into the database when an error was discovered either because the error was recognized on subsequent imaging, in consultation with clinicians, or during interdisciplinary conferences. All errors had been

reviewed and verified by consensus at the quarterly QA conference. Surgical findings or biopsies were available to confirm the error for 28% of cases.

Major errors were defined as those that a general practice radiologist would not be expected to make, and one that altered management in a way potentially detrimental to the patient (Figs. 1 and 2). At the quarterly QA conference, there was far greater agreement on what constituted a serious or major missed finding (which had material clinical consequences) than there was for a minor missed finding (which did not). Therefore, this study focused on major errors only. Those abdomen/pelvis CT studies (henceforth, only “abdominal CT”) containing at least one major error from a 10-year period formed the basis of this study.

Errors were classified in accord with ACR guidelines, specifically as errors of perception, interpretation, and communication [7]. False positive perceptual errors occurred when normal findings are interpreted as pathologic. False negative perceptual errors occurred when pathologic findings were missed and unreported. Errors of interpretation occurred when pathologic findings were identified but were misattributed, misperceived, or misdiagnosed [7]. Errors of communication occurred when an

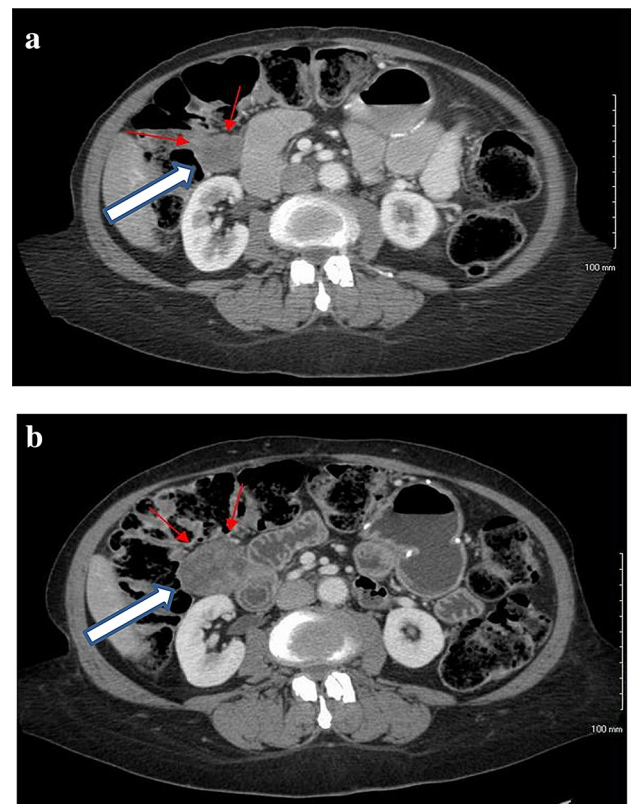


Fig. 1 **a, b** A late morning false negative error (“miss”) in a 62 years old female with pancreatic cancer. **a** Mesenteric metastasis not identified at 11:02 am on a Thursday. **b** Interval growth of the mass 4 months later



Fig. 2 **a, b** A last hour afternoon false negative error in an 81 years old female with cervical cancer. **a** Left pelvic wall nodal metastasis not seen at 4:09 pm on a Monday. **b** Interval growth of the mass 2 months later

important finding was not effectively communicated to the caring provider.

The identified studies were categorized by the day of the week and hour of the day that the study was interpreted. The hour of interpretation was acquired from the time stamp on the dictation, and assigned within the hour the interpretation was made (e.g. a studies read at 11: 15 and 11: 45 would both be assigned to the 11 am hour). Therefore, the data from the weekend days were not considered comparable to those of the weekdays, and excluded from analysis. Weekend cases were read separately from the weekday cases, and not included in weekday case collection. Study volume data over this same period were also obtained by day of the week and time of day, so to normalize the data based on case volume (Table 1). Within the time period of the study, a total of 483,481 abdominal CT studies were read. The distribution of cases by day is given in Table 1. For each case identified, data were also collected with regard to anatomic location of the error, study indication, study size (number of images), and patient category (inpatient vs. outpatient).

At our academic institution, all cases were interpreted by a subspecialty trained abdominal radiologist during the work week, which began at 7:00 am and ended at 5:00 pm from Monday to Friday. Outside of these hours, initial interpretation was rendered by a resident, which was reviewed at 7:00–7:30 am the next day by staff. Through the night hours (5 pm to 7 pm), subspecialty radiologists were available to provide consultation and answer questions as needed. The hour of interpretation was assigned according to the time stamp on the resident dictation record, and therefore reflects resident error rather than specialist error.

We tested for non-random variation in the daily composition of cases by comparing study size (number of images), study indication, and error location. For the distribution of errors across study indication (cancer, infection, vascular/trauma/other), a Chi-square test (with eight degrees of freedom) was calculated between the 5 weekdays. The average number of images per study was also calculated and compared across the 5 weekdays using the Kruskal–Wallis rank sum test. Further, the data were dichotomized into 5-year periods and tested for consistency over time for error rate by weekday, study indication, and anatomic site of the error using chi-square tests.

Standard errors of the volume-adjusted error rates were obtained based on the binomial distribution. The null hypothesis of constant error rates over time was tested using a weighted logistic regression model with linear time as predictor. Error rates at different time points were compared using two-sample *t*-test based on normal approximation to the binomial distribution. All analyses performed using R version 4.0.2. All statistical tests were considered statistically significant at *p* values less than 0.05.

Results

A total of 252 major errors were identified from the QA database. The majority of errors ($n=211$, 84%) were false negative perceptual errors (Figs. 3 and 4). A smaller number ($n=41$, 16%) were interpretive errors. Tests for data consistency over time showed no significant variation from the first half of the study to the second for errors per weekday ($p=0.67$), study indication ($p=0.19$), or anatomic site of error ($p=0.60$).

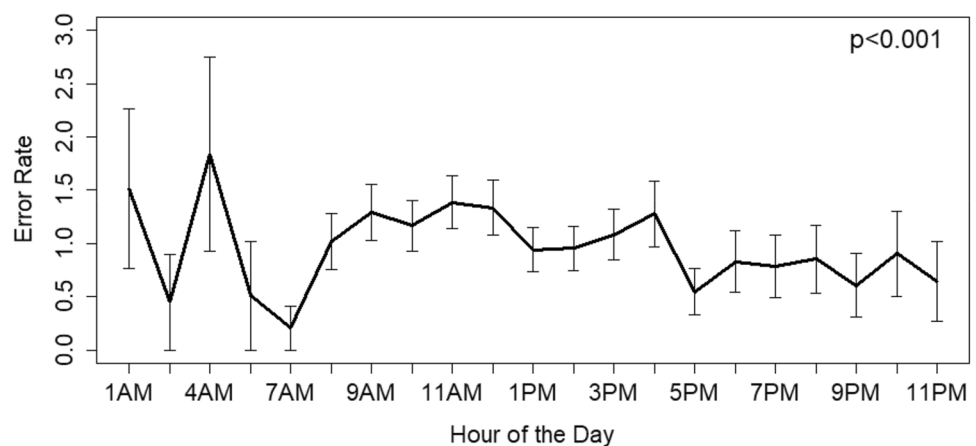
More errors were made on Monday than on any other day of the week ($n=58$). The majority of errors occurred in outpatients ($n=186$) and the most common indication was neoplasm follow-up ($n=176$). The missed finding was related to the indication for the study 71% of the time ($n=179$).

The most common anatomic regions of error were the hepatobiliary system ($n=44$, 17%), mesentery ($n=35$, 14%). Errors were also frequent in the vasculature ($n=30$, 12%), body wall ($n=27$, 11%), bowel ($n=27$, 11%), and pancreas

Table 1 Study volume by hour and day of the week

Hour	Sunday	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Totals
12 AM	976	914	831	783	776	914	824	6018
1 AM	893	709	624	656	650	693	743	4968
2 AM	875	642	581	580	577	602	765	4622
3 AM	808	534	482	497	580	480	775	4156
4 AM	801	485	430	477	513	533	841	4080
5 AM	723	470	419	465	479	544	767	3867
6 AM	645	464	442	473	502	508	647	3681
7 AM	562	1595	1651	1606	1607	1653	508	9182
8 AM	708	4893	4954	4823	4752	4991	611	25,732
9 AM	753	6303	5970	6143	5878	6145	710	31,902
10 AM	1018	6751	6726	6704	6553	6567	1027	35,346
11 AM	1214	7297	6909	7166	7217	6714	1305	37,822
12 PM	1378	6366	6052	6346	6181	5942	1423	33,688
1 PM	1447	7124	6967	7076	7161	6454	1580	37,809
2 PM	1497	6837	6863	6929	7080	6491	1619	37,316
3 PM	1497	6224	5825	5795	6014	5981	1486	32,822
4 PM	1415	4318	4198	3825	4130	3997	1523	23,406
5 PM	1431	3775	3600	3222	3660	3354	1523	20,565
6 PM	1332	3252	3092	2967	3108	2991	1350	18,092
7 PM	1241	2957	2816	2625	2881	2754	1356	16,630
8 PM	1243	2805	2717	2460	2511	2327	1345	15,408
9 PM	1111	2085	1990	2041	1974	1847	1260	12,308
10 PM	1061	1787	1589	1686	1688	1382	1150	10,343
11 PM	1053	1309	1311	1280	1339	1291	1135	8718
Totals	25,682	79,896	77,039	76,625	77,811	75,155	26,273	438,481

Fig. 3 Diurnal variation of error rates during weekdays. The error rate has been normalized to case volume for each hour. Note rising rates of error as the morning progresses, and near the end of the workday (5:00 pm). Error bars indicate one standard error. Of note, studies acquired from 5 pm to 7 am were categorized by the time of interpretation by the covering resident, and those studies from 7 am to 5 pm were categorized by the time of interpretation by the staff radiologist



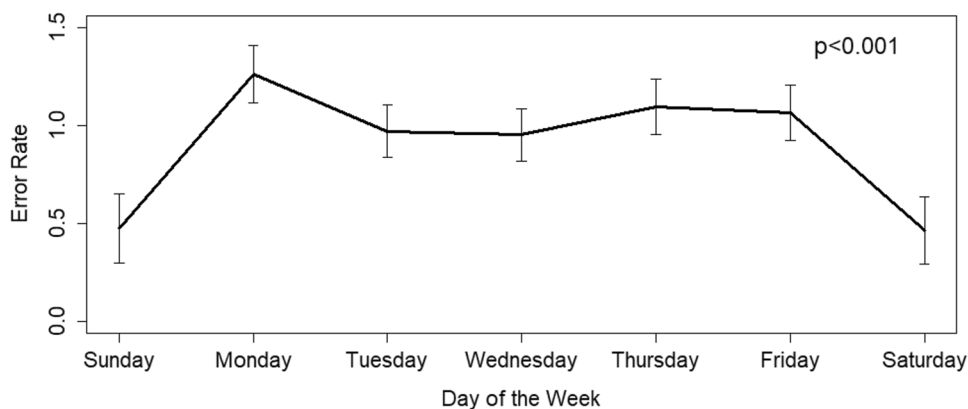
($n = 10, 4\%$). Two errors contributed to patient demise: both involving bowel infarction/necrosis.

Major error rates normalized by study volume increased through the mid to late morning (9am to 12 pm) ($p < 0.001$ for trend test over 7am–12 pm), and then decreased progressively through the afternoon until 4 pm ($p = 0.1152$ for trend test over 12 pm–4 pm), when a rise in the error rate was seen (Figs. 3 and 4). Overall tests of time-constancy of error rates by day and hour were statistically significant

(both p -values < 0.001). Error rates and error by volume rates were highest on Mondays and lowest on Saturdays and Sundays by a factor of two ($p < 0.001$).

No significant difference was found between the 5 weekdays for the distribution of errors across diagnostic categories (cancer, infection, vascular/trauma/other) ($p = 0.42$). Further, the average number of axial images per study was similar for weekdays: Monday (256), Tuesday (213), Wednesday (292), Thursday (273), Friday (325). No

Fig. 4 Error rate normalized to case volume for days of the week. Error bars indicate one standard error. Of note, studies acquired on Sunday and Saturday were initially interpreted by the covering resident



significant difference was found in the average number of axial images per study across weekdays ($p = 0.21$).

Discussion

This study demonstrates a discernable pattern of diurnal variation in major error rates in the interpretation of abdominal CT studies. During the workweek, the error rates were highest in the late morning and at the close of the workday. These rates were greatest on Mondays.

The data on weekend and night hour error rates are more problematic and were therefore not considered in the analysis. First, the volume of cases is less, and the possibility of chance variation is greater. Second, the reading of these cases was performed largely by residents of varying levels of experience, and therefore the reading group by hour is different from the reading group by hour during the weekdays (subspecialty trained abdominal radiologists). Finally, cases were read as a batch by the abdominal imaging faculty either the next morning or in the evening, rather than as an ongoing, continuous process during the weekday. Weekend cases were not included with those of weekdays (note the similar number of cases at 7 am for all weekdays in Table 1).

A potential problem with our data could be that there is a non-random variation in the daily composition of cases, such that certain types of cases (either more complex or less) might tend to occur on a particular day because certain types of clinics (pancreatic cancer clinic, for example) meet on particular days. Though this is certainly possible, there was relative constancy of error through the weekdays (except for Monday) as well as relative constancy in diagnostic category (cancer, infection, vascular, etc.), and location of error. Indeed, there were no statistically significant differences by weekday for the average number of images in the studies per day, the distribution of error by anatomic region, or the distribution of diagnostic category by day. These considerations indicate that daily variations in the complexity of cases is unlikely to have a substantial effect on our results. Finally,

the length of time over which the sample was drawn would also tend to mitigate non-random effects.

There is a small literature on diurnal variation of radiologist performance [3–6, 8]. To date, studies have focused on almost exclusively on plain film interpretation (mammograms and chest radiographs) and many were conducted outside a radiology reading room environment and before the advent of the modern PACS workstations [3–5]. Further, these studies have included readers of a wide range of experience, from medical students to radiology residents to expert subspecialty radiologists, and included errors of varying clinical significance.

In medical endeavors outside of radiology, diurnal variation of human performance has been verified in a range of tasks [6, 10]. Adverse medical decision-making, for instance, have shown time of day variation for events such as unintentional injuries, diagnosis delays, and misidentification of a body part. The time of day where clinical performance was poorest was the mid-afternoon [6, 10].

Our study suggests that radiologists experience periods through the day where physical, psychological, mental, and cognitive resources wax and wane. Factors contributing to the depletion could include fatigue, eye-strain, stress, haste, hunger, and metabolism [9, 11, 12]. Fatigue, as a specific factor of study, have been demonstrated to result in cognitive errors performance in mammography and other types of radiology interpretation [9, 11–14]. That said, the increased error rate on Mondays suggests that additional factors may be present.

Equally important considerations are the factors that might contribute to the restoration of energies, such as breaks, food, exercise, micro-naps, and sunlight. Performance reading mammograms, for examples, seems to vary with time of day, time spent working, and break scheduling [11]. Scheduling adjustments, such as half-day shifts on the most taxing services, might be instituted to mitigate exhaustion of energies.

That error would be greatest on Mondays is consistent with studies indicating a decline in mental acuity after

periods of relative inactivity [15–18]. Some have suggested that radiologists should rapidly read through a practice image set on a daily or weekly basis, so to fine-tune perception and combat inattention [15–18].

This study has some limitations. First, the diurnal variation observed in this study may not be present in practices where workday schedules are structured differently. Second, we did not determine whether error rate was affected by the time from study acquisition to final dictation, particularly for those weekend cases that were given a preliminary interpretation at night and a final interpretation in the morning. Another limitation of this study is the likelihood of incomplete capture of cases containing error. It is impossible to insure that all cases containing error will be discovered and reported. Indeed, individual error can be repeated and perpetuated over one or more subsequent studies [11]. That said, requiring that collected were major errors, which is to say highly consequential errors, would tend to make these errors more likely to be discovered over time. Further, this requirement would also tend to mitigate hindsight bias and inter-rater variability. That said, this constraint also resulted in a relatively small sample size by day and by time of day, and may have excluded cases that others would have considered clinically significant.

The decision as to what constituted a major error was necessarily subjective, though we found greater agreement for errors of this category than for errors that were “likely” or “possibly” clinically relevant. Hence the narrowing of the focus of the study. Patient records were not reviewed to determine specifically if management was in fact altered when the error was discovered. Arguably, non-major errors could be a proxy for errors in general. It was difficult to determine, however, whether a non-major finding was missed or simply dismissed by the reader as not sufficiently important to report.

The focus of the study was on major errors made by the faculty radiologists. We did not control for the involvement of residents or fellow. Except for the weekend and overnight cases, which were always initially read by a trainee, we did not identify those cases that had been previously reviewed by a trainee (double-read) and those that had not. Nor did we account for overlapping shifts, lunch coverage, staffing levels, clinic schedules, study priority (stat or routine), or proximity to vacations. Finally, we did we compare individual radiologists as to error rates, case volumes, or absences.

This study demonstrates that error rates in abdominal CT do appear to vary with time of day and day of the week. During the workweek, error rates were highest in the late morning and at the close of the workday, and greater on Mondays than other days. These variations could indicate the effects of fatigue, waning attention, or haste. Radiologists need to be cognizant of these patterns, so to find means to remain vigilant in the late morning and close of day.

References

1. Rutenfranz J, Colquhoun WP. Circadian rhythms in human performance. *Scand J Work Environ Health* 1979. 5:167e77
2. Colquhoun WP. Circadian variations in mental efficiency. W.P. Colquhoun (Ed.), *Biological rhythms and human performance*, Academic Press, London (1971), pp. 39–107
3. Gale AG, Murray D, Millar K, Worthington BS: Circadian variation in radiology. In: Gale AG, Johnson F Eds. *Theoretical and applied aspects of eye movement research*. Elsevier, London, 1984.
4. Taylor-Phillips S, Clarke A, Wallis M, et.al. The time course of cancer detection performance. *Proc SPIE Med Imag* 2011 7966:796605–1–8.
5. Al-s’adi M, McEntee MF, Ryan E. Time of day does not affect radiologists’ accuracy in breast lesion detection. *Proc SPIE Med Imag* 2011. 7966:796608–1–7.
6. Alshabibi AS, Suleiman ME, Tapia KA, Brennan PC. Effects of time of day on radiological interpretation. *Clin Radiol*. 2020. 75(2):148–155.
7. Goldberg-Stein S, Frigini LA, Long S, et al. ACR RADPEER Committee White Paper with 2016 Updates: Revised Scoring System, New Classifications, Self-Review, and Subspecialized Reports. *J Am Coll Radiol* 2017. 14(8):1080-1086.
8. Hallas P, Ellingsen T. Errors in fracture diagnoses in the emergency department--characteristics of patients and diurnal variation. *BMC Emerg Med*. 2006;6:4. Published 2006 Feb 16.
9. Brogdon BG, Kelsey CA, Moseley RD: Effect of fatigue and alcohol on observer perception. *Am J Roentgenol* 1978. 130:971–974.
10. Buckley, D., Reyment, J. and Curtis, P. "The witching time: diurnal patterns in adverse events of clinical management", *Clinical Governance: An International Journal* 2009. Vol. 14 No. 4, pp. 281-288
11. Krupinski, EA, Reiner, BI. Real-Time Occupational Stress and Fatigue Measurement in Medical Imaging Practice. *J Digit Imaging* 2012. 25, 319–324
12. Krupinski EA, Berbaum KS, Caldwell RT, et al. Long radiology workdays reduce detection and accommodation accuracy. *J Am Coll Radiol* 2010, pp. 698–704,
13. Stec N, Arje D, Moody AR, et al. A systematic review of fatigue in radiology: is it a problem? *AJR* 2018, 210, pp. 799-806,
14. Krupinski EA: Reader fatigue interpreting mammograms. *Lect Notes Comput Sci* 2010. 6136:312–318.
15. Waite S, Grigorian A, Alexander RG, et.al. (2019) Analysis of Perceptual Expertise in Radiology – Current Knowledge and a New Perspective. *Front Hum Neurosci* 2019. 13:213.
16. Reicher MA, Wolfe JM. . Let’s Use Cognitive Science to Create Collaborative Workstations. *Journal of the American College of Radiology* 2016. 13(5):571-575.
17. Waite S, Farooq Z, Grigorian A, et. al. A Review of Perceptual Expertise in Radiology-How it develops, How we can test it, and Why humans still matter in the era of Artificial Intelligence. *Academic Radiology* 2020. 27 (1), pp. 26-38.
18. Tajmir SH, Alkasab TK. Toward Augmented Radiologists: Changes in Radiology Education in the Era of Machine Learning and Artificial Intelligence. *Acad Radiol* 2018. 25 (6), pp. 747-750.

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.